Ann. Inst. Statist. Math.
Vol. 47, No. 3, 461–464 (1995)

QUASI PROFILE AND DIRECTED LIKELIHOODS FROM ESTIMATING FUNCTIONS

O. E. BARNDORFF-NIELSEN

Department of Theoretical Statistics, Institute of Mathematics, Aarhus University, Ny Munkegade, DK-8000 Aarhus C, Denmark

(Received May 16, 1994; revised January 6, 1995)

Abstract. Frequently, corresponding to a given estimating equation it would be desirable to have a scalar combinant having parametric derivative equal to the estimating function since such a combinant may serve as a quasi log likelihood. In general this cannot be achieved but it is nevertheless possible to define a quasi profile log likelihood and also a quasi directed likelihood, for an arbitrary one-dimensional parameter of interest and with the standard kind of distributional limit behaviour.

Key words and phrases: Directed likelihood, estimating equations, profile likelihood.

1. Introduction

Let $g = g(y; \theta)$ be an unbiased estimating function for a d-dimensional parameter θ. Here y denotes the full data set and g is a vector of dimension d. The estimate of θ determined by g will be denoted by $\hat{\theta}$, i.e.

$$g(y; \hat{\theta}) = 0. \tag{1.1}$$

Occasionally we shall write g_θ for the vector g.

In case g_θ is the score vector of the log likelihood function l for data y, so that g_θ is of the form

$$g_\theta = l_{/\theta}, \tag{1.2}$$

where / indicates differentiation, we have that

$$\mathrm{cov}\{g_\theta\} = -E\{g_{\theta/\theta}\}. \tag{1.3}$$

In general, the relation (1.3) does not hold. However, as noted by McCullagh ((1991), Section 11.7), we can achieve that it does, without changing the properties of the estimator $\hat{\theta}$, by multiplying g by a matrix that may depend on θ but is constant in y. We assume henceforth that this has been done.

Even when (1.3) holds there does not in general, when $d > 1$, exist a scalar function $h = h(y;\theta)$ whose gradient with respect to θ equals g, i.e. which has the property that

$$h_{/\theta} = g_\theta. \tag{1.4}$$

In case such a function h is available it may be thought of as a log likelihood-like function for θ and it may be used, in analogy with ordinary log likelihoods, for setting quasi likelihood regions, for constructing quasi profile likelihoods, etc.

Essentially, whether (1.3) holds or not, the necessary and sufficient condition for the existence of a h satisfying (1.4) is that the matrix $g_{\theta/\theta}$ be symmetric or, equivalently and in the language of differential geometry, that the vector field given by g_θ is conservative. When g_θ is conservative the function h is said to be the potential function associated to g_θ.

The desirability of existence of a potential function, or 'quasi likelihood', has been stressed by Li and McCullagh (1994) in the context of generalized linear models with g as the quasi score

$$g(y;\theta) = \{\dot{\mu}(\theta)\}^\top \{V(\theta)\}^{-1}\{y - u(\theta)\},$$

in the standard notation. As a remedy when g is not conservative they suggest a method of projecting the quasi score onto a class of conservative estimating functions and they show that the projected quasi score has many properties similar to those of an ordinary log likelihood function. Alternative approaches, also based on projection, have been proposed by McLeish and Small (1992) and Li (1993).

As will be discussed in the following, the problem of nonexistence of a primitive h of g may usually be circumvented as regards one-dimensional parameters of interest.

Suppose that θ is partitioned as $\theta = (\psi, \chi)$ into a one-dimensional parameter of interest ψ and a $(d-1)$-dimensional nuisance parameter χ, and that g_θ is similarly divided as (g_ψ, g_χ) with g_χ being the estimating function for χ that would be used also if ψ was known. The estimate for χ derived from g_χ when ψ is considered as known will be denoted by $\hat{\chi}_\psi$, i.e.

$$g_\chi(\psi, \hat{\chi}_\psi) = 0. \tag{1.5}$$

We will indicate that a combinant (i.e. a function of the data y and the parameter θ) $q = q(y;\theta)$ is evaluated at $\hat{\theta}$ or $(\psi, \hat{\chi}_\psi)$ by the symbols ^ and ~, respectively. Thus $\hat{q} = q(y;\hat{\theta})$ and $\tilde{q} = q(y;\psi,\hat{\chi}_\psi)$. By convention, the operations ^ and ~ are taken to be always the last carried out; without this convention a symbol such as $\tilde{l}_{/\psi\psi}$ would be ambiguous.

The fact on which we will draw is that if there is a log likelihood function l for θ and if we take $g = l_{/\theta}$ then the profile log likelihood $l_P = l_P(\psi) = l(\psi, \hat{\chi}_\psi)$ for ψ satisfies

$$l_{P/\psi} = \tilde{l}_{/\psi}. \tag{1.6}$$

In more explicit notation equation (1.6) is

$$\partial l(\psi, \hat{\chi}_\psi)/\partial\psi = \partial l(\psi, \chi)/\partial\psi \,|_{\chi \to \hat{\chi}_\psi} .$$

This implies that the normalized profile log likelihood for ψ may be written as

$$\bar{l}_P = l_P(\psi) - l_P(\hat{\psi}) = \int_{\hat{\psi}}^{\psi} \tilde{l}_{/\psi} d\psi. \tag{1.7}$$

In view of this we propose, for an arbitrary estimating function g as specified above, to define the quasi profile score $\tilde{g}_\psi = g_\psi(\psi, \hat{\chi}_\psi)$ and a corresponding quasi profile log likelihood $h_P = h_P(\psi)$ by

$$h_P(\psi) = \int g_\psi(\psi, \hat{\chi}_\psi) d\psi. \tag{1.8}$$

This may then be used for setting quasi likelihood intervals for ψ, for defining a quasi directed likelihood $t = t_\psi$ for ψ, etc. In Section 2 we discuss the relevant properties of $h_P(\psi)$ and t.

The normalized quasi profile log likelihood $\bar{h}_P = \bar{h}_P(\psi)$ is given by

$$\begin{aligned} \bar{h}_P(\psi) &= h_P(\psi) - h_P(\hat{\psi}) \\ &= \int_{\hat{\psi}}^{\psi} g_\psi(\psi, \hat{\chi}_\psi) d\psi. \end{aligned} \tag{1.9}$$

In case the interest parameter ψ is multidimensional it may happen, though this will be a rare situation, that the vector field determined by the quasi profile score $\tilde{g}_\psi$ is conservative. Equation (1.8) then determines a unique (up to an additive constant) potential function for $\tilde{g}_\psi$. Note in this connection that the relation (1.6) is valid also for multivariate ψ. The potential function may sensibly be used as a quasi profile log likelihood for ψ. Firth and Harris (1991) suggested this in the context of quasi likelihood for generalized linear models. (However, the example they consider, which is for a multiplicative random effects model, is in error. Apparently, what has been overlooked is that $\tilde{g}_\psi = g(\psi, \hat{\chi}_\psi)$ depends on ψ both 'directly' and through $\hat{\chi}_\psi$.)

2. Distribution properties of $h_P(\chi)$ and t

The quasi directed likelihood $t = t_\psi$ is defined as

$$t = \operatorname{sign}(\hat{\psi} - \psi)\{-2\bar{h}_P(\psi)\}^{1/2}. \tag{2.1}$$

Under regularity conditions of standard type the statistic t will be approximately standard normally distributed and $-2\bar{h}_P(\psi)$ will be approximately χ^2-distributed on one degree of freedom.

To see this note first that, under such conditions, the partial estimator $\hat{\chi}_\psi$ is approximately normal with mean χ and covariance matrix

$$\operatorname{cov}\{\hat{\chi}_\psi\} \doteq b^{\chi,\chi^\top} \operatorname{cov}\{g_\chi\} b^{\chi,\chi} \tag{2.2}$$

where

$$b^{\chi,\chi} = E\{g_{\chi/\chi}\}^{-1}. \tag{2.3}$$

Denoting coordinates of χ by $\chi_i, \chi_j, \ldots$, the matrix $g_{\chi/\chi}$ has (i,j)-th element

$$g_{\chi_j/\chi_i} = \partial g_{\chi_j}/\partial \chi_i$$

and the inverse matrix $b^{\chi,\chi} = [b^{j,i}]$ is not in general symmetric.

Combining (1.3) and (2.2) we have

$$\operatorname{cov}\{\hat{\chi}_\psi\} \doteq -b^{\chi,\chi^{\top}} E\{g_{\chi/\chi}\} b^{\chi,\chi}. \tag{2.4}$$

Taylor expansion of $h_P(\psi)$ around $\hat{\psi}$ yields

$$-2\bar{h}_P(\psi) \doteq (\hat{\psi}-\psi)^2\{\tilde{g}_{\psi/\psi} - \tilde{g}_{\chi/\psi}\tilde{b}^{\chi,\chi}\tilde{g}_{\psi/\chi}\} \tag{2.5}$$

where we have used the fact that

$$\hat{\chi}_{\psi/\psi} = -\tilde{g}_{\chi/\psi}\tilde{b}^{\chi,\chi}, \tag{2.6}$$

as follows on differentiating the equation (1.5), determining $\hat{\chi}_\psi$, with respect to ψ.

Further, writing I for $\operatorname{cov}\{g_\theta\}$ we have approximately, in view of (1.3),

$$\begin{aligned} -\{\tilde{g}_{\psi/\psi} - \tilde{g}_{\chi/\psi}\tilde{b}^{\chi,\chi}\tilde{g}_{\psi/\chi}\} &\doteq I_{\psi\psi} - I_{\psi\chi}I_{\chi\chi}^{-1}I_{\chi\psi} \\ &= \{I^{\psi\psi}\}^{-1} \end{aligned} \tag{2.7}$$

in standard notation for block matrices. Since $\hat{\psi} - \psi \sim N(0, i^{\psi\psi})$ the above-mentioned conclusions concerning the asymptotic distributions of $-2\bar{h}_P(\psi)$ and t follow.

Acknowledgements

Helpful comments from referees of the paper are gratefully acknowledged.

References

Firth, D. and Harris, I. R. (1991). Quasi-likelihood for multiplicative random effects, *Biometrika*, **78**, 545–555.

Li, B. (1993). A deviance function for the quasi-likelihood method, *Biometrika*, **80**, 741–753.

Li, B. and McCullagh, P. (1994). Potential functions and conservative estimating functions, *Ann. Statist.*, **22**, 340–356.

McCullagh, P. (1991). Quasi-likelihood and estimating equations, *Statistical Theory and Modelling* (eds. D. V. Hinkley, N. Reid and E. J. Snell), Chapter 11, Chapman and Hall, London.

McLeish, D. L. and Small, C. G. (1992). A projected likelihood function for semiparametric models, *Biometrika*, **79**, 93–102.